Economics Letters 123 (2014) 348–351

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

A Monte Carlo study of a factor analytical method for fixed-effects dynamic panel models

Milda Norkute *

Department of Economics, Lund University, P.O. Box 7082, S-220 07 Lund, Sweden

HIGHLIGHTS

- A new factor analytical method to estimate fixed-effects dynamic panel data models is considered.
- The method is proposed by Bai (2013a) and it has the feature that it is asymptotically bias free.
- We provide Monte Carlo evidence of the good small-sample performance of this method.
- Our results thus complement Bai's theoretical study.

ARTICLE INFO

Article history:
Received 31 January 2014
Accepted 26 March 2014
Available online 1 April 2014

JEL classification:
C13
C33

Keywords:
Dynamic panel data models
Heteroscedasticity
Monte Carlo simulations

ABSTRACT

In a recent article Bai (2013a) proposes a new factor analytical method (FAM) for the estimation of fixed-effects dynamic panel data models, which has the unique and very useful property that it is asymptotically bias free. In this paper we provide Monte Carlo evidence of the good small-sample performance of FAM, that complement Bai's theoretical study.

1. Introduction

It is well known that the within-groups (WG) estimator of the autoregressive coefficient in dynamic panel data models is subject to an incidental parameter bias, typically referred to as the "Nickell bias". This bias, which occurs because the number of fixed-effects parameters grows without bound, is of order $1/T$, indicating that the WG estimator is inconsistent in panels where T is small even if N goes to infinity (see, for example Baltagi, 2008). This problem has led to increased interest in generalized method of moments (GMM), see Baltagi (2008) for an overview of this literature. However, although estimation methods based on GMM will successfully remove the incidental parameter bias, they are instead biased of the order $1/N$. This means that they are inconsistent in panels where N is small even if T goes to infinity (see Alvarez and Arellano, 2003). GMM approaches are also known to suffer from problems of small-sample inefficiency (see, for example Kiviet, 1995), and weak instrumentation (see Roodman, 2009). Another possibility is to use bias correction methods, which have the advantage of not being reliant on instrumental variables (see, for example Kiviet, 1995; Hahn and Kuersteiner, 2002). These methods are, however, still biased in panels where T is small due to the approximation error in the asymptotic bias term. To the best of our knowledge, FAM is the only existing estimation method for dynamic panel data models that is bias free under a wide range of conditions (see, for example Moon et al., forthcoming).

In a recent paper, Bai (2013a,b) proposes a factor analytical method (FAM) to estimate fixed-effects dynamic panel data models. One of the main features of this method is that there is no need for consistent estimation of the fixed-effects themselves but only their variance, which means that with this approach there is no "Nickell bias". Indeed, as Bai (2013a,b) shows one can even allow for heteroscedasticity and still there is no bias. In contrast to other approaches, FAM therefore allows for asymptotically unbiased inference regardless of whether the incidental parameters are in the mean or in the variance. Of course, as is well known, asymptotic results need not provide accurate approximations in small samples. Thus, while certainly very promising, the usefulness of FAM from an applied point of view is yet to be proven since its small-sample

* Tel.: +46 46 222 7919; fax: +46 462224118.
E-mail address: Milda.Norkute@nek.lu.se.

http://dx.doi.org/10.1016/j.econlet.2014.03.020
0165-1765/

properties are still unknown. The aim of the present paper is to fill this gap in the literature.

2. FAM

The data generating process (DGP) considered in the present paper is the same as in Bai (2013a), and is given by

$$y_{it} = \rho y_{it-1} + \mu_i + \delta_t + \varepsilon_{it}, \quad (1)$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, μ_i and δ_t are individual- and time-specific fixed-effects, respectively, $y_{10} = \cdots = y_{N0} = 0$, $|\rho| < 1$, and ε_{it} is an error term that is assumed to be independently distributed with $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma_t^2 > 0$ and $E(\varepsilon_{it}^4) < \infty$. It is further assumed that $S_\mu = (N-1)^{-1}\sum_{i=1}^N(\mu_i - \bar{\mu})^2 > 0$, where $\bar{\mu} = N^{-1}\sum_{i=1}^N \mu_i$. Eq. (1) can be written in a matrix form as

$$y_i = \Gamma 1_T \mu_i + \Gamma\delta + \Gamma\varepsilon_i, \quad (2)$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$, $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, $\delta = (\delta_1, \ldots, \delta_T)'$ and $1_T = (1, \ldots, 1)'$ are all $T \times 1$ vectors. The matrix Γ is $T \times T$ and is given by

$$\Gamma = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ \rho & 1 & 0 & \ldots & 0 \\ \rho^2 & \rho & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{T-1} & \ldots & \rho^2 & \rho & 1 \end{bmatrix}.$$

The sample covariance matrix of y_i is given by $S_y = (N-1)^{-1}\sum_{i=1}^N(y_i - \bar{y})(y_i - \bar{y})'$, where $\bar{y} = N^{-1}\sum_{i=1}^N y_i$. Under the above assumptions, it can be shown that

$$E(S_y) = \Sigma(\theta) = \Gamma(1_T 1_T' S_\mu + \Psi)\Gamma', \quad (3)$$

where $\Psi = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_T^2)$ and $\theta = (S_\mu, \rho, \sigma_1^2, \ldots, \sigma_T^2)'$ is the vector containing the parameters of interest. The model in (2) can be seen as a common factor model with factor loading $\Gamma 1_T$ and score μ_i, suggesting that the estimation can be carried out using methods designed for such models (see, for example Anderson and Amemiya, 1988). FAM is based on quasi-maximum likelihood whereby θ is estimated by minimizing the following "discrepancy function"

$$Q(\theta) = \log(|\Sigma(\theta)|) + \mathrm{tr}\,(S_y \Sigma(\theta)^{-1}). \quad (4)$$

Denote by $\hat{\rho}$ the resulting estimator of ρ. As Bai (2013a, Theorem 1) shows, as $N, T \to \infty$ with $NT^{-3} \to 0$,

$$\sqrt{NT}(\hat{\rho} - \rho) \to_d N(0, \gamma^{-1}), \quad (5)$$

where $\to_d$ signifies convergence in distribution and $\gamma = \lim_{T\to\infty} T^{-1}\sum_{t=2}^T \sigma_t^{-2}(\sigma_{t-1}^2 + \rho^2\sigma_{t-2}^2 + \cdots + \rho^{2(t-2)}\sigma_1^2)$.[1] Hence, under the above conditions, there is no asymptotic bias and the estimator is asymptotically efficient. Moreover, the condition $NT^{-3} \to 0$ is not necessary if we are only concerned with consistency (this condition is only imposed to ensure a simple form for the limiting distribution). In fact, consistency only requires $N \to \infty$.

Remark 1. Note that θ only contains S_μ, not $\mu_1, \ldots, \mu_N$, and also how the time-specific fixed-effects are removed by subtracting $\bar{y}$ in S_y. This means that the incidental parameter problem caused by the growing dimension of θ does not arise. The way in which the incidental parameter problem is treated in FAM is therefore very different from the conventional approach of either performing the within transformation or by taking first differences. Of course, the dimension of θ is still growing in T; however, the estimation of $\sigma_1^2, \ldots, \sigma_T^2$ does not affect the consistency of $\hat{\rho}$ (see Bai, 2013a, for an explanation).

Remark 2. While under the above assumptions ε_{it} is homoskedastic in i, this is not necessary. If ε_{it} is heteroskedastic in both i and t, then the estimation can proceed in exactly the same way as in the above, but then $\sigma_1^2, \ldots, \sigma_T^2$ only capture the average variances over cross-sections (and not the variances themselves). This removes the incidental parameter problem, as the dimension of θ does not depend on N.

3. Monte Carlo results

The DGP used in this section is given by (1) with $\varepsilon_{it} \sim N(0, \sigma_{it}^2)$, $\mu_i \sim U(1, 2)$ and $\rho \in \{0, 0.5, 0.95\}$. We run four distinct experiments[2]:

A. $\sigma_{it}^2 = 1$, $\delta_t = 0$;
B. $\sigma_{it}^2 = 1$ if $i < \lfloor N/2 \rfloor$ and $\sigma_{it}^2 = 2$ otherwise, $\delta_t = 0$;
C. $\sigma_{it}^2 = 1$ if $t < \lfloor T/2 \rfloor$ and $\sigma_{it}^2 = 1/3$ otherwise, $\delta_t = 0$;
D. $\sigma_{it}^2 = 1$, $\delta_t \sim U(1, 2)$.

In experiment A ε_{it} is homoscedastic, whereas in experiments B and C heteroscedasticity is permitted by allowing for two distinct variance regimes. In experiments A–C the time-specific effects are absent, while in experiment D this is no longer the case. In each experiment the data are generated for 5000 panels with $T \in \{5, 10, 50\}$ and $N \in \{10, 50, 100\}$.

The small-sample performance of FAM is compared to the performance of four other estimators (of ρ); WG, the bias-corrected ordinary least squares (OLS) estimator of Hahn and Kuersteiner (2002), the Anderson and Hsiao (1981) instrumental variables (IV) estimator using lagged levels as instruments,[3] and the GMM estimator of Arellano and Bond (1991), where the last three are henceforth denoted by bcOLS, AHl and abGMM, respectively. A large number of results were produced, but due to space constraints we focus here on the bias and root mean squared error (RMSE). Estimation using abGMM is computationally intensive, so for this estimator we only report results for the case when $T = 5$.

The results for experiment A are presented in Table 1. We see that the bias of FAM is close to zero for all the sample sizes considered. In fact, in terms of bias FAM outperforms the other estimators. We also see that the bias becomes smaller in absolute value as T and N increases, a finding that is in line with the $\sqrt{NT}$-consistency of FAM (irrespective of the relative rate of expansion of N and T). As expected, with T fixed, WG is seriously biased and there is no improvement as N increases. abGMM is also noticeable biased when N is small; however, the performance improves as N increases. Similarly, although severely biased when N and T are small, the performance of bcOLS improves when T increases. We also see that the performance of bcOLS is much worse when $\rho = 0.95$ than for $\rho = 0.5$ or $\rho = 0$, which is in agreement with the findings in the previous literature. The performance of AHl is quite good and is only dominated by that of FAM.

FAM is superior, not only in terms of bias, but also in terms of RMSE. The RMSE of FAM is decreasing in both T and N. The fact that RMSE is also decreasing in ρ is in agreement with the theoretical result that the asymptotic variance of FAM is inversely related to the absolute value of ρ. One can also observe, that except for the case when $\rho = 0.95$, the RMSE of bcOLS and FAM are quite comparable, which is consistent with the fact that the both estimators are asymptotically efficient. The results provided in Table 1 further suggest that the least efficient estimator is AHl, which exhibits the highest RMSE.

[1] Under homoscedasticity, $\gamma = (1 - \rho^2)^{-1}$ (see Bai, 2013b, Theorem S.2).

[2] All computational work is performed in GAUSS 11 and the BFGS algorithm is used for constrained optimization with non-negativity constraints imposed on the variance parameters.

[3] We use level rather than first-differenced instruments, as the IV estimator based on the latter instruments has a singularity point and exhibits high variance over a wide range of parameter values (see Arellano, 1989).